**程序描述**

**一、业务背景介绍**

某电信公司希望了解影响客户流失的主要因素，提取了在指定期限内流失客户与留存客户的100条数据进行生存回归分析。

**二、数据结构**

|  |  |  |
| --- | --- | --- |
| 字段名 | 数据类型 | 含义 |
| churn | boolean | 截止观测期，客户的留存状态（1=流失，0=留存） |
| tenure | integer | 截止观察期，客户接受服务的时间 |
| age | integer | 客户年龄 |
| Address | integer | 客户居住区域 |
| income | integer | 客户收入 |
| ed | integer | 客户的教育水平，1=“高中以下”；2=“高中”；3=“专科”；4=“本科”；5=“研究生” |
| employ | integer | 客户目前工作的在任市场 |
| gender | boolean | 客户性别 |

**三、伪代码**

data=读取用户数据

data[‘status’]转换为布尔类型

绘制时间与生存概率的步阶图

添加x轴、y轴标签

显示图片

estimator=生存分析基本模型

data\_x=自变量

data\_ytp=应变量数据类型

data\_yy=应变量

data\_y=应变量转换数据类型

模型训练（data\_x，data\_y）

输出自变量与相应回归系数